Multisource Image Fusion Algorithm Based On A New Evidential Reasoning Approach

Mickaël Germain[1], Matthieu Voorons[1], Jean-Marc Boucher[2], Goze B. Bénié[1] et Éric Beaudry[1]

(1) Centre d'applications et de recherches en télédétection (CARTEL),
Université de Sherbrooke, Canada, Email : mickael.germain@usherbrooke.ca
tél. : (1-819)821-7180
(2) École Nationale Supérieure des Télécommunications de Bretagne,
France, Email : jm.boucher@enst-bretagne.fr
tél. : 33 (0)2 29 00 11 11

Abstract— **We propose a new way to initialize the mass function of the Dempster-Shafer theory of evidence. The new initialization process is based on a fuzzy statistical approach and uses the FSEM algorithm (Fuzzy Statistical Estimation Maximization). This allows to classify image in "pure" and "fuzzy" regions, and thus enable an optimal estimation of the inaccuracy and uncertainty of the classification.**

We apply our new evidential reasoning approach for the fusion of a Landsat multispectral image with vegetation indices and a digital elevation model.

Keywords— **Remote Sensing, Classification Algorithm, Fusion Algorithm, Evidential Reasoning, Fuzzy Logic**

INTRODUCTION

This article describes a new data fusion algorithm which is a part of the SITI project (Intelligent System of Image Processing). The purpose of this project is to design and develop new algorithms for analysing, segmenting and extracting information based on an expert fusion process of optical, radar or auxiliary data.

Data fusion related to a same object or a same scene becomes more and more essential in remote sensing applications. It is often necessary to associate additional and/or redundant information, in order to reject, confirm or create a decision. A definition of data fusion was formulated by Bloch and Maître [3] : "data fusion is the joint use of heterogeneous information for the assistance with the decision-making". This definition emphasizes the essential points of a fusion process :

- the heterogeneity of the data makes it possible to provide additional information for sources of similar or different nature ;

- the joint use of information enables to specify the importance of the final decision. Indeed, if a decision is made for each kind of data separately, then the process can not be considered as a fusion process anymore ;

- the goal of fusion is to provide an aid in the decision-making process.

There are mainly three models of fusion operators cited in the scientific literature : probabilistic bayesian models , fuzzy models and models resulting from the Dempster-Shafer theory of evidence.

The probabilistic bayesian models are the most cited models ; the concept of fusion is deduced from the Bayes rule. However, in the bayesian models there is a confusion between two antagonist concepts : the uncertainty and the inaccuracy. Moreover, we have to note that the performances of the bayesian data fusion tend to be decrease when the number of information sources increases.

One of the most known non-probabilistic techniques is the fuzzy theory. This technique, introduced by Zadeh [13], represents information in the form of explicit functions of membership. The disadvantage of the fuzzy theory is that it characterizes the uncertainty in an implicit way, only the inaccurate property of information is represented [3].

The Demspter-Shafer (DS) theory of evidence allows to represent at the same time the inaccuracy and uncertainty using confidence, plausibility and credibility functions. It defines a framework of understanding representing all the subsets of the classes space. The principal advantage of this theory is to affect a degree of confidence which is called *mass function* to all simple and composed classes, and to take into account the ignorance of the information. However, there is no generic method to define the mass functions. Most of the time, they are computed using an empirical method which depend on the nature of the information. Thus, we will present, in the next sections, a new global solution with a more rigorous way to deal with the concepts of uncertainty and inaccuracy in the DS theory.

THE DEMPSTER-SHAFER THEORY OF EVIDENCE

The DS theory of evidence was first introduced by Dempster [6] and formalized by Shafer [11]. This mathematical theory is composed of three distinct parts : the definition of the mass functions, the combination process and the decision-making.

The definition of the mass definition

A mass function can be compared with a degree of confidence one can have in the studied data. It have to be set between values 0 and 1, where 1 stands for a total confidence and 0 for no confidence at all. In the terminology of Dempster and Shafer, we do not define anymore data or classes, but only "hypotheses". Then, a mass function will be defined on a hypotheses set, called the *frame of discernment*. It represents a set of mutually exclusive and exhaustive propositions.

Let us note the hypotheses set Θ composed of single mutually exclusive subset θ_i. The DS fusion works on a single hypothesis, but it works also on all subset composed of several single hypotheses. So the DS fusion process is based on 2^Θ elements called *propositions*.

A mass function for one source and for one proposition is defined as follows :

$$m : 2^\Theta \rightarrow [0, 1] \tag{1}$$

$$\sum_{A \in 2^\Theta} m(A) = 1 \tag{2}$$

$$m(\phi) = 0 \tag{3}$$

By using this representation model one can assign a confidence value to a set of composed hypotheses. This value shows that it is impossible to dissociate the set of assigned hypotheses. It is the main advantage but also the principal difficulty of the DS method. Indeed, there is no generic method to define a mass value on a single or a composed hypothesis.

Evidence combination

The greatest advantage of DS theory is the robustness of its way of combining information coming from various sources with the DS orthogonal rule. For instance, let us denote two mass distributions m_1 and m_2 from two sources. Then, the DS combination can be represented by the following orthogonal rule :

$$(m_1 \oplus m_2)(A) = \frac{\sum\limits_{B_1 \cap B_2 = A} m_1(B_1) m_2(B_2)}{1-K}, \quad K \neq 1 \tag{4}$$

$$K = \sum_{B_1 \cap B_2 = \phi} m_1(B_1) m_2(B_2) \tag{5}$$

K is considered as a normalization factor and is interpreted as a measure of conflict between the various sources. In addition, it is a representation of the empty set mass function. Thus, the larger K, the more the sources conflict and the less sense their combination has. If $K = 1$, then the sources are totally contradictory.

Decision making

Unlike the bayesian theory, where the decision criterion is often the maximum of likelihood, the DS theory gives many solutions to take a decision. The are several ways to decide which is the most reliable hypothesis from single or unions of propositions.

The decision making rules that are the generally used are : maximum of belief, maximum of plausibility or compromises between them. We can find an exhaustive list of decision criterion in [4].

MASS FUNCTION DEFINITION

Initialization methods for the mass function in the DS theory are various and depend on the considered application framework considered. According to the method applied for the initialization process, we can set with a correctly way the mass functions were selected, it is possible to more or less correctly translate the various aspects of uncertainty and the inaccuracy. There are currently two main categories of applications for the mass function initialization. The first one is based on probabilistic methods and lead to method like the consonant, partially consonant or disonant distributions [4][8]. The drawback of these methods is that they are defined in an empirical way for the composed propositions. They do not take into account fuzzy data. The second mass function initialization category is based on fuzzy analysis [2][12]. These techniques use the membership functions as mass function, but they do not respect highly incertain information as the presence of noise on image for example. The originality of this project is to use a fuzzy statistical algorithm based on the FSEM (Fuzzy Stochastic Estimation Maximization) in order to better characterize the concepts of uncertainty and inaccuracy in the mass function definition.

FUZZY STATISTICAL CLASSIFICATION METHOD

Principle of the fuzzy SEM algorithm

Classification processes of remote sensing images do not represent the complex reality of a studied area. In fact, let us consider the problem of segmenting a satellite image into two classes : “water” and “vegetation”. There may be some pixels with only vegetation or water, but others, as in a boggy area, in which water and vegetation are simultaneously present. In the first case, the pixel will be called a *pure* pixel and in the second case, it will be called a *mixed* pixel.

Some people proposed solutions to resolve this problem, Caillol *et al.*[5] introduced fuzzy data in some classical statical models. To counter this drawback, some algorithms introduce a fuzzy version of statistical modelling. Thus, parameter estimation stochastic algorithms have been modified to take fuzzy data into account such as the Expectation Maximisation (EM), Stochastic Expectation Maximization(SEM) and Iterative Conditional Estimation (ICE) algorithm. All theses algorithms only apply for the case of a maximum of two pure classes. Estimation algorithms become too complex if the number of pure classes increases.

In the following section, we propose to generalize the Fuzzy SEM (FSEM) for K classes with the following hypothesis : a fuzzy class or a mixed class cannot be composed of more than two "pure" classes [7]. This hypothesis is widely observed in pratice, like “water-vegetation”, “trees-house” and “agricultural-vegetation” fuzzy classes, whereas “trees-houses-agricultural” mixed classes or others are fairly unusual.

A complete description of the FSEM algorithm can be find in [7]. Let us define the unobservable random field $X = (X_s)_{s \in S}$ taking its values in a finite set of classes Θ. We denote $Y = (Y_s)_{s \in S}$ the observed random field which is a corrupted version of X, and S be the set of pixels $S = \{1, ..., n\}$.

Let us define two independant gaussian random variables y_i and y_j associated with the two pure classes θ_i and θ_j. The gaussian densities f_i and f_j are defined by the distributions $\aleph(m_i, \sigma_i)$ and $\aleph(m_j, \sigma_j)$ respectively. The fuzzy density defined for a fuzzy class between θ_i and θ_j can be simply obtained by the following linear relation [10] :

$$y_s = \varepsilon y_i + (1-\varepsilon) y_j, \quad \varepsilon \in]0,1[_{ij} \tag{6}$$

where ε is the mixture coefficient between θ_i and θ_j.

In this context, we show that the fuzzy distribution is a gaussian density $\aleph(m_{ij}(\varepsilon), \sigma_{ij}(\varepsilon))$:

$$m_{ij}(\varepsilon) = (1-\varepsilon) m_i + \varepsilon m_j \tag{7}$$

$$\sigma_{ij}^2(\varepsilon) = (1-\varepsilon)^2 \sigma_i^2 + \varepsilon^2 \sigma_j^2 \tag{8}$$

Finally, the pure distribution density is defined by :

$$f_i(y) = \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-\frac{(y-m_i)^2}{2\sigma_i^2}} \tag{9}$$

and the fuzzy distribution density :

$$f_{ij}(\varepsilon, y) = \frac{1}{\sqrt{2\pi\sigma_{ij}^2(\varepsilon)}} e^{-\frac{(y-m_{ij}(\varepsilon))^2}{2\sigma_{ij}^2(\varepsilon)}} \quad (10)$$

Thus, by generalizing the approach defined in Caillol *et al.* [5], the pixel density is given by the following formula :

$$p(y_s) = \sum_{i=1}^{K} \pi_i f_i(y_s) + \sum_{i=1}^{K-1} \sum_{j=i+1}^{K} \int_0^1 \pi_{ij} f_{ij}(\varepsilon, y_s) d\varepsilon \quad (11)$$

where π_i and π_{ij} correspond to, respectively, the *a priori* probability of the pure and the fuzzy classes.

With the bayesian theory, we can describe the *a posteriori* probability for each class :

- for the set of pure classes :

$$P(x_s = \theta_i / Y_s = y_s) = \frac{\pi_i f_i(y_s)}{p(y_s)} \quad (12)$$

- for the set of fuzzy classes :

$$P(x_s = \theta_{f_{ij}} / Y_s = y_s) = \frac{\int_0^1 \pi_{ij} f(\varepsilon, y_s) d\varepsilon}{p(y_s)} \quad (13)$$

From the definition of the above probabilities we can estimate the unknown parameters : $\pi_i, \pi_j, \pi_{ij}, m_i, m_j, \sigma_i, \sigma_j$ from a sample of X. A full description of the FSEM algorithm can be found in [7].

FSEM applied to multispectral data

The fuzzy statistical analysis described previously is defined for only one spectral band. However the data on which we work are are made of several, say N, bands. Those N images have to be analysed, so in this context the unobservable random field can ne represented by $X^N = (X_s^N)_{s \in S}$.

The introduction of the multidimensionnal property of the data increases the algorithmic complexity of the problem. The generalization of the equations (12,13) to N bands is not trivial. The solution is to set a simplifying hypothesis in order to make the algorithm practically realizable. The most used hypothesis is *the conditional independence* which stipulates that, knowing the class θ_i, the joint density of two variables y_s^1 and y_s^2 is the product of the densities of each variable :

$$f_i(y_s^1, y_s^2) = f_i(y_s^1).f_i(y_s^2) \quad (14)$$

This hypothesis can be reinforce by applying a principal component analysis (PCA) which reduces the inter-band correlation and decrease the number of spectral bands. The equation (11) is written for a number N of spectral bands in the following way :

$$p(y_s^N) = \sum_{i=1}^{K} \pi_i f_i(y_s^N) + \sum_{i=1}^{K-1} \sum_{j=i+1}^{K} \int_0^1 \pi_{ij} f_{ij}(\varepsilon, y_s^N) d\varepsilon \quad (15)$$

The simplification of equation 15 with the independence conditional hypothesis result in the following formula :

$$p(y_s^N) = \sum_{i=1}^{K} \pi_i \prod_{n=1}^{N} f_i(y_s) + \sum_{i=1}^{K-1} \sum_{j=i+1}^{K} \int_0^1 \pi_{ij} \prod_{n=1}^{N} f_{ij}(\varepsilon, y_s) d\varepsilon \quad (16)$$

Fuzzy statistical method for mass function initialization

Mathematically, we can define non-normalized masses for all the simple et composed hypotheses as follows :

$$\tilde{m}(\theta_i) = \prod_{n=1}^{N} f_i(y_s) \quad (17)$$

$$\tilde{m}(\theta_i \cup \theta_j) = \int_0^1 \prod_{n=1}^{N} f_{ij}(\epsilon, y_s) d\epsilon \quad (18)$$

where $f_i(y_s)$ and $f_{ij}(\epsilon, y_s)$ are the conditional densities described in the previous section.

Application to remote sensing images

Data set

The studied zone is the region of “Grand Lake”, located in the area of Gooze Bay, Labrador. It is mainly composed of different forest densities and clear cuts. The last cartographic update realized in 1988.

We use the PCA on the LANDSAT image to reduce the data to three bands containing 95% of the information. We also compute the Tasseled cap images, to extract the brightness, greenness and wetness informations, see figure 1.b to 1.d. We also have auxiliary information relative to the altitude of the studied area (fig.1.e). All this complementary and redundant information have to be extracted in a rigorous way and the orthogonal sum of DS is used to combine them.

The extraction of the information is carried out by the FSEM algorithm from PCA data and Tasseled Cap images. The pure and fuzzy densities extracted are used in the initialization process to compute simple and composed hypotheses of the evidential theory. We apply a sober filter on the altitude information, it results in a map containing the slope information of the area. We use this information to initialize the mass functions according to a slope threshold above which confidence for simple and composed hypotheses “water” or “boggy” is weak.

Results analysis

A simple probabilistic unsupervised classification based on the SEM algorithm gives 55 % of good classification (fig.2.a). The contribution of DS fusion initialized by the FSEM algorithm with the Tasseled Cap transformation can improve the classification quality in particular for boggy and vegetation classes. The rate of classification is 61 % (fig.2.b). The contribution of slope information (fig.2.c) removes some natural artefact related to LANDSAT TM data acquisition. In fact, the shadow of some clouds is classified like “water” or “boggy” with the SEM algorithm. Some pixels classified as water or boggy are on high slope areas which is not realistic. Those pixels are in fact cloud shadows. The DS fusion removes this kind of error by decreasing the credibility associated to this decision. The rate of classification after DS fusion is 63 % .

The final step is based on an evidential markovian approach. Each classified pixel is studied with a spatial context and its own mass function after the DS fusion [1]. The algorithm used in this step is the evidential ICM (Iterative Conditional Mode). The rate of good classification after this algorithm increase to 66 % (fig.2.d).

CONCLUSION

The new evidential fusion approach can distinct in a better way the uncertainty et inaccuracy notion in the mass functions. The FSEM algorithm compared to others algorithms used in the DS fusion enables possible to distribute a more realistic density for each simple or composed hypothesis. We have seen that the mass function initialization proposed in this article ease the fusion process. The mass function are no more estimated using an empirical approach, so the algorithm is completely automatic and we do not need any a priori information about the data.

Application to remote sensing multispectral images with some ecological indices and auxiliary data related to slope information give a better classification result on the final decision-making. Redundant and heterogeneity information decrease some ambiguities related to a lack of data and some artefacts. Thus, the DS fusion developed method improves the result in comparaison with the result of LANDSAT TM SEM classification.

REFERENCES

[1] Bendjebbour, A., Delignon, Y., Fouque, L., Samson, V. and Pieczynski, W. (2001) Multisensor Image Segmentation Using Dempster-Shafer Fusion in Markov Fields Context, IEEE Trans. On Geoscience and Remote Sensing, vol. 39, p. 1789-1798.

[2] Bentabet, L., Zhu, Y. M., Kaftdandjian, V., Babot, D. and Rombaut, M.(2000) Use of fuzzy clustering for determining mass functions in Dempster-Shafer theory, presented at Signal Processing Proceedings, WCCC-ICSP 2000. 5th International Conference on.

[3] Bloch, I. and Maître, H. (1994) Fusion de Données en Traitement d'Images : modèles d'information et décisions, Traitement du Signal, vol. 6, p. 1811-1823.

[4] Bracker, H. (1996) Utilisation de la théorie de dempster/Shafer pour la classification d'images satellitaires à l'aide de données multi-sources et multi-temporelles. : Rennes I, p. 178.

[5] Caillol, H., A., H. and Pieczynski, W. (1993) Fuzzy random fields and unsupervised image segmentation, IEEE Transaction on Geoscience and Remote Sensing, vol. 31, p. 801-810.

[6] Dempster, A. P. (1968) A generalisation of Bayesian inference, Journal of the Royal Statistical Society, p. 205-247.

[7] Germain, M., Voorons, M., Boucher, J.-M. and Benie, G. B.(2002) Fuzzy statistical classification method for multiband image, presented at Proceedings of the Fifth International Conference on Information Fusion, Annapolis, Maryland, USA.

[8] Lee, T., Richards, J. A. et Swain, P. H. (1987) Probabilistic and Evidential Approches for Multisource data Analysis, IEEE Transactions on Geoscience and Remote Sensing, vol. 25, p. 283-293.

[9] Peddle, D. R. (1993) An Empirical Comparaison of Evidential reasoning, Linear Discriminant Analysis and Maximum Likelihood Algorithms for Alpine Land Cover Classification, Canadian Journal of Remote Sensing, vol. 19, p. 31-44.

[10] Salzenstein, F. and Pieczynski, W. (1997) Parameter estimation in hidden fuzzy Markov random fields and image segmentation, Graphical Models and Image Processing, vol. 59, p. 205-220.

[11] Shafer, G. (1976) A Mathematical "Theory of Evidence", vol. Princeton, NJ.

[12] Verikas, A., Malmqvist, K. and Bacauskiene, M.(2000) Combining Neural Networks, Fuzzy Sets, and Evidence Theory Based Approaches for Analysing Colour Images, presented at Neural Networks, 2000. IJCNN 2000, Proceedings of the IEEE-INNS-ENNS International Joint Conference on.

[13] Zadeh, L. A. (1965) Fuzzy sets, Information and Control, vol. 8, p. 338-352.

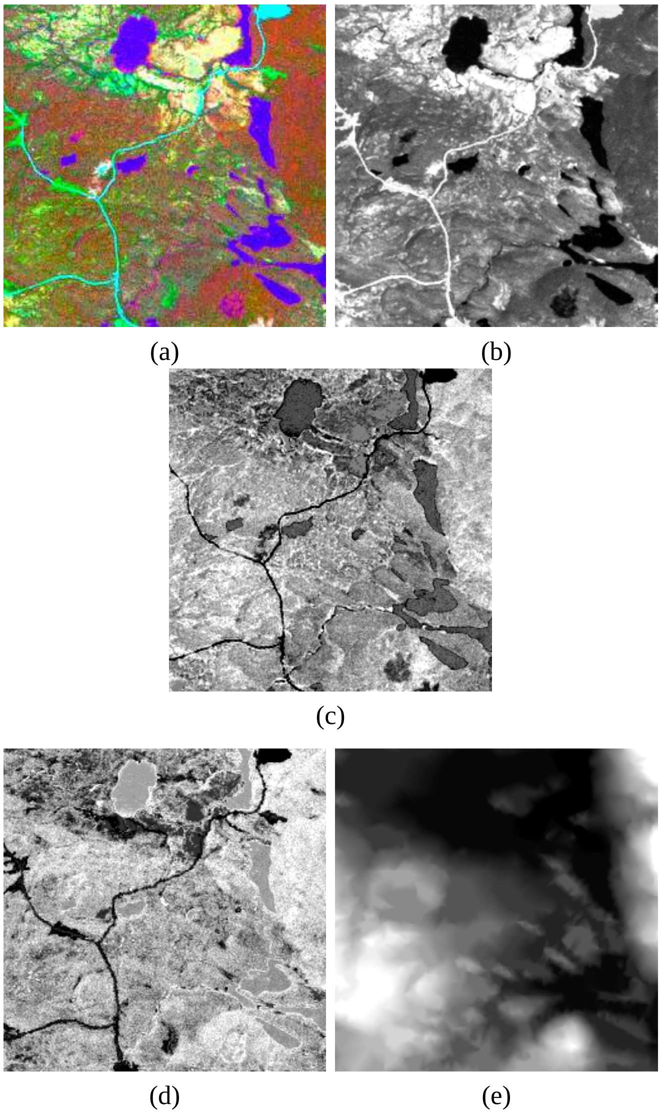

Fig. 1. (a) image composed of 3 principal components from LANDSAT TM multispectral images, (b) to (d) Brightness, Greenness, and Wetness Tasseled Cap indices respectively, (e) Altitude information

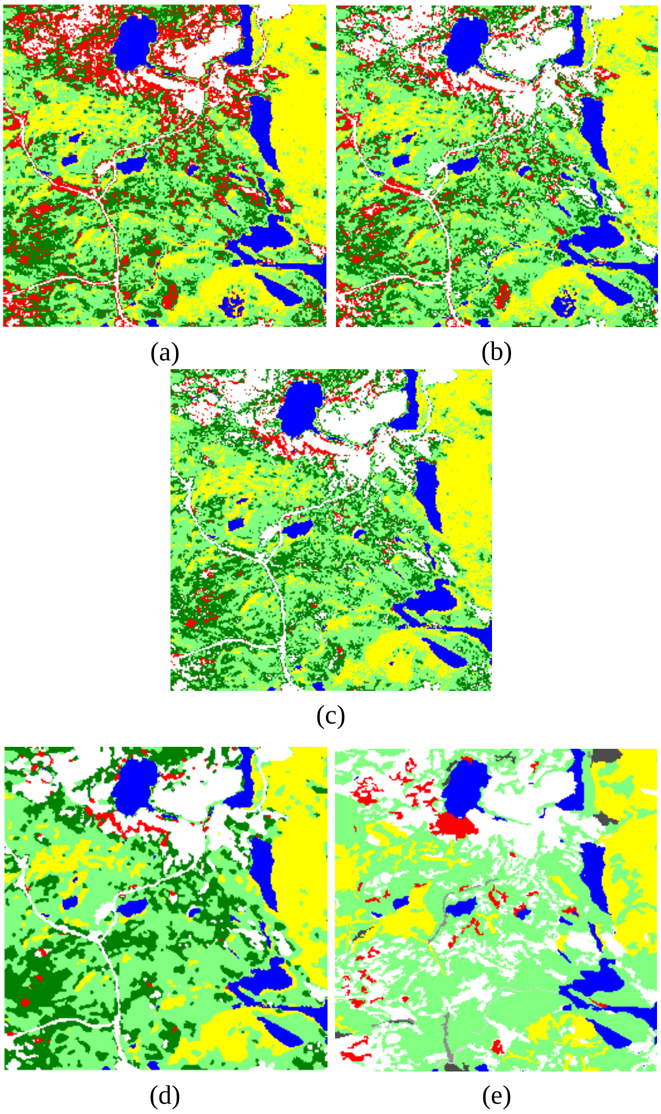

Fig. 2. (a) SEM, (b) FSEM + Tasseled Cap fusion, (c) FSEM + Tasseled Cap fusion + slope, (d) Evidential SEM, (e) Ground Truth